

Distributed Bayesian Detection Under Unknown Observation Statistics

Xiaojing Shen, Member, IEEE, Pramod K. Varshney, Fellow, IEEE, Yunmin Zhu ^{*†}

September 20, 2012

Abstract

In this paper, distributed Bayesian detection problems with unknown prior probabilities of hypotheses are considered. The sensors obtain observations which are conditionally dependent across sensors and their probability density functions (pdf) are not exactly known. The observations are quantized and are sent to the fusion center. The fusion center fuses the current quantized observations and makes a final decision. It also designs (updated) quantizers to be used at the sensors and the fusion rule based on all previous quantized observations. Information regarding updated quantizers is sent back to the sensors for use at the next time. In this paper, the conditional joint pdf is represented in a parametric form by using the copula framework. The unknown parameters include dependence parameters and marginal parameters. Maximum likelihood estimation (MLE) with feedback based on quantized data is proposed to estimate the unknown parameters. These estimates are iteratively used to refine the quantizers and the fusion rule to improve distributed detection performance by using feedback. Numerical examples show that the new detection method based on MLE with feedback is much better than the usual detection method based on the assumption of conditionally independent observations.

keywords: Bayesian detection, copula-based dependence modeling, copulas, maximum likelihood estimation, distributed detection, information fusion

1 Introduction

Distributed detection has received considerable attention over the last few decades [1, 2, 3, 4, 5]. The Bayesian formulation of distributed detection was first considered by Tenney and Sandell [6] for parallel

^{*}This work was supported in part by U.S. Air Force Office of Scientific Research (AFOSR) under Grants FA9550-10-1-0263 and FA9550-10-1-0458 and in part by the NNSF of China (# 61004138 and 61273074).

[†]Xiaojing Shen and Pramod K. Varshney (corresponding author, varshney@syr.edu) are with the Department of Electrical Engineering and Computer Science, Syracuse University, NY, 13244, USA. Yunmin Zhu is with Department of Mathematics, Sichuan University, Chengdu, Sichuan 610064, China. Xiaojing Shen (shenxj@scu.edu.cn) is on leave from Department of Mathematics, Sichuan University, Chengdu, Sichuan 610064, China.

sensor network structures. For conditionally independent sensor observations, they proved that the optimal decision rules at the sensors are likelihood ratio (LR) quantizers. The optimal thresholds to quantize LR at individual sensors can be determined by solving a set of coupled nonlinear equations. When the quantizers are fixed, Chair and Varshney in [7] derived an optimum fusion rule, once again based on the LR test. Over the past several years, many excellent results on distributed detection based on the assumption of conditionally independent sensor observations have been derived that are available in [1] and references therein. The emerging wireless sensor networks [8] motivated the optimality of LR quantizers to be extended to non-ideal detection systems where sensor outputs are to be communicated through noisy, possibly coupled channels to the fusion center [9, 10].

When sensor observations are dependent, Tsitsiklis and Athans [11] provided a rigorous mathematical analysis demonstrating the computational difficulty in obtaining the optimum quantizers. Some progress has been made for the dependent observations case (see [12, 13, 14, 15]). For example, difficulties encountered when dealing with dependent observations were discussed in [14]. In [16], for distributed dependent observations and a fixed fusion rule, the authors proposed a computationally efficient iterative algorithm for computing a discrete approximation of the optimal quantizers. The finite-step convergence of this algorithm was proved. By combining the methods proposed in [7] and [16], an efficient algorithm to simultaneously search for the optimal fusion rule and the optimal quantizers was derived in [17]. Recently, the authors of [18] introduced a new framework for distributed detection with conditionally dependent observations. The new framework can identify several classes of problems with dependent observations whose optimal quantizers resemble the ones for the independent case. In addition, copula-based distributed Neyman-Pearson detection and hypothesis testing using heterogeneous dependent data have been proposed in [19, 20]. The copula based approach provides a systematic and elegant approach to characterize dependence and obtain decision rules at the sensors and the fusion center.

In all previous studies on distributed Bayesian detection with dependent observations [11, 12, 13, 14, 15, 16, 17, 18], the conditionally dependent joint pdfs of the sensor observations are assumed known. When the dependence among sensors is unknown, the usual approach is to ignore dependence and assume that the sensor observations are independent. The focus of this paper is distributed Bayesian detection in the context of unknown conditionally dependent pdfs. We also assume that prior probabilities of the hypotheses are unknown. The specific scenario (see Figure 1) is that the sensors obtain observations which are conditionally dependent across sensors. The observations are quantized and are sent to the fusion center. The fusion center fuses the current quantized observations and makes a final decision. It designs (updated) quantizers to be used at the sensors and the fusion rule based on all previous quantized observations. Information regarding updated quantizers is sent back to the sensors for use at the next time.

In [17], we have presented an iterative algorithm to design the quantizers at the sensors and the fusion center for Bayesian distributed detection. In this paper, we extend the work in [17] to the situation when the prior probabilities and the joint pdf of conditionally dependent observations are not known. The conditional joint pdf is represented in a parametric form by using the copula framework. The unknown parameters

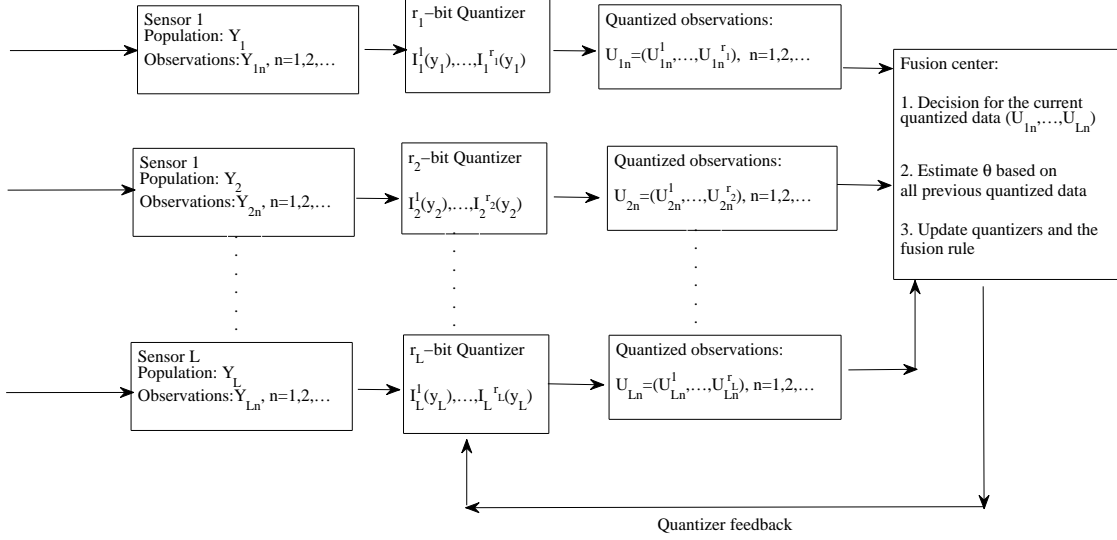


Figure 1: Distributed detection system with feedback of quantizer information

include dependence parameters and marginal parameters (parameters corresponding to marginal pdfs). MLE with feedback based on quantized data is proposed to estimate the unknown parameters. Its asymptotic efficiency can be guaranteed by employing the result that we have developed in [21] with an asymptotic variance which is equal to the inverse of a convex linear combination of Fisher information matrices based on J groups of different feedback quantizers. These estimates are iteratively used to refine the quantizers and the fusion rule to improve distributed detection performance by using feedback. Numerical examples show that the new detection method based on MLE with feedback is much better than the usual detection method based on the assumption of conditionally independent observations. Better detection performance can be obtained by increasing the number of feedbacks and the number of observations during each estimation step.

The rest of the paper is organized as follows. Problem formulation is given in Section 2. In Section 3, copula-based parametric pdfs are constructed. MLE with feedback based on quantized data is also proposed. In Section 4, an efficient distributed detection algorithm with unknown pdfs and unknown joint prior probabilities is presented based on quantized observations and different updated quantizers. In Section 5, numerical examples are given that exhibit the superior performance of our approach. In Section 6, concluding remarks are provided.

2 Problem Formulation

An L -sensor Bayes detection system with two hypotheses H_0 and H_1 is considered. A parallel architecture with feedback is assumed (see Figure 1). Each sensor acquires observations $Y_i, i = 1, \dots, L$ whose dimension is assumed to be one for notational simplicity in this paper. The case of high dimensions can be similarly considered. The i -th sensor quantizes the observation vector to r_i bits ($r_i \geq 1$) by r_i measurable indicator quantization functions:

$$I_i^1(y_i) : y_i \in \mathbb{R} \rightarrow \{0, 1\}; \dots; I_i^{r_i}(y_i) : y_i \in \mathbb{R} \rightarrow \{0, 1\}, \quad (1)$$

for $i = 1, \dots, L$. Here, each binary quantizer $I_i^t(y_i)$ partitions the space into two regions that could be continuous regions or union of discontinuous regions. Moreover, we denote the quantization functions by

$$I(y|r) \triangleq (I_1(y_1)', \dots, I_L(y_L)')' \in \mathbb{R}^r, \quad (2)$$

where

$$I_i(y_i) \triangleq (I_i^1(y_i), \dots, I_i^{r_i}(y_i))', \quad i = 1, \dots, L, \quad (3)$$

and $r = \sum_{i=1}^L r_i$ is the total number of bits available to transmit observations from the sensors to the fusion center. Once the r_i -bit binary quantized measurements $I_i(Y_i)$ are generated at sensor $i, i = 1, \dots, L$, they are transmitted to the fusion center. The fusion center then makes a final decision 0/1 based on $I_i(Y_i), i = 1, \dots, L$ using a fusion function $F(u_1, \dots, u_L), u_i = 0/1, i = 1, \dots, L$, i.e.,

$$F(I_1(Y_1), \dots, I_L(Y_L)) = 0/1. \quad (4)$$

If the prior probabilities of hypotheses and the conditional pdfs $p(y_1, \dots, y_L|H_j), j = 0, 1$ are known, the goal of the distributed Bayesian detection system is to design a set of optimal sensor quantizers $I_1(y_1), \dots, I_L(y_L)$ and an optimum fusion rule $F(u_1, \dots, u_L)$ such that the following Bayes cost functional is as small as possible.

$$\begin{aligned} & C(I_1(y_1), \dots, I_L(y_L); F) \\ & \triangleq C_{00}P_0P(F(I_1(Y_1), \dots, I_L(Y_L)) = 0|H_0) + C_{01}P_1P(F(I_1(Y_1), \dots, I_L(Y_L)) = 0|H_1) \\ & \quad + C_{10}P_0P(F(I_1(Y_1), \dots, I_L(Y_L)) = 1|H_0) + C_{11}P_1P(F(I_1(Y_1), \dots, I_L(Y_L)) = 1|H_1), \end{aligned} \quad (5)$$

where C_{ij} are cost coefficients; P_0 and P_1 are the prior probabilities for the hypotheses H_0 and H_1 ; $P(F(I_1(Y_1), \dots, I_L(Y_L)) = i|H_j)$ is the probability that the fusion center decides in favor of hypothesis i given hypothesis $H_j, i, j = 0, 1$, which can be computed based on $p(y_1, \dots, y_L|H_j), j = 0, 1$ (see, e.g., [1]).

However, in distributed detection systems with limited bandwidth, the joint conditional pdfs are hard to obtain by traditional pdf estimation methods based on raw observations. In some situations, the prior probabilities of H_0 and H_1 are also unknown. Quantized observations and initial quantizers is all the information

available at the fusion center. The specific scenario considered here is that the sensors obtain observations which are conditionally dependent across sensors. The observations are quantized and are sent to the fusion center. The fusion center fuses the current quantized observations and makes a final decision. It also designs (updated) quantizers to be used at the sensors and the fusion rule based on all previous quantized observations. Information regarding updated quantizers is sent back to the sensors for use at the next time.

In summary, when the prior probabilities and conditionally dependent pdfs are unknown, the fusion center faces the problem of how to make a decision for the current set of quantized observations and the problem of how to improve the detection performance by using all previous quantized observations in time.

The first problem is how to design the quantizers and the fusion rule, which is a static problem when the prior probabilities and the conditional pdfs are known. Previously developed methods (see, e.g., [7, 15, 16, 17, 18]) can be used to solve the problem. Thus, we concentrate on how to update/estimate the unknown prior probabilities and conditionally dependent pdfs using all the previous quantized data. Note that the problem relies on the updated quantizers that are fed back to the sensors, which results in quantized observations not having identical distributions temporally.

3 Copula-based Maximum Likelihood Estimation with Feedback

3.1 Copula-based dependence modeling

In distributed detection with dependent observations, the performance of the detection system depends on the exploitation of dependence among sensor measurements. In most previous works, when dependence is not known, independence is usually assumed across sensors for simplicity. Here, we will model dependence by parametric copulas. Actually, copula is a distribution function whose one-dimensional marginals are uniform.

Lemma 3.1. (*Sklar's Theorem, see [22] or [23]*) *Let F be an L -dimensional distribution function with marginals F_1, F_2, \dots, F_L . Then there exists an L -copula $C(v_1, v_2, \dots, v_L)$, $v_i \in [0, 1], i = 1, \dots, L$ such that for all (y_1, \dots, y_L) in \mathbb{R}^L ,*

$$F(y_1, y_2, \dots, y_L) = C(F_1(y_1), F_2(y_2), \dots, F_L(y_L)). \quad (6)$$

If F_1, F_2, \dots, F_L are all continuous, then C is unique; otherwise, C is uniquely determined on $\text{Ran } F_1 \times \text{Ran } F_2 \times \dots \times \text{Ran } F_L$. Conversely, if C is an L -copula and F_1, F_2, \dots, F_L are distribution functions, then the function F defined by (6) is an L -dimensional distribution function with marginals F_1, F_2, \dots, F_L .

From Sklar's Theorem, the joint pdf is equivalent to

$$p(y_1, y_2, \dots, y_L) = c(F_1(y_1), F_2(y_2), \dots, F_L(y_L)) \prod_{i=1}^L p_i(y_i), \quad (7)$$

where $p_i(y_i)$ and $F_i(y_i)$, $i = 1, \dots, L$ are marginal pdfs and distribution functions of continuous random variables respectively; $c(v_1, v_2, \dots, v_L)$, $v_i \in [0, 1]$, $i = 1, \dots, L$ is the copula density function,

$$c(v_1, v_2, \dots, v_L) = \frac{\partial^L C(v_1, v_2, \dots, v_L)}{\partial v_1, \dots, \partial v_L}. \quad (8)$$

If measurements are conditionally independent, then $p(y_1, y_2, \dots, y_L) = \prod_{i=1}^L p_i(y_i)$ and $c(v_1, v_2, \dots, v_L) \equiv 1$. When the measurements are dependent, $c(v_1, v_2, \dots, v_L) \neq 1$ and all dependence information among measurements is contained in $c(v_1, v_2, \dots, v_L)$. Thus, copula framework allows us flexibly represent the dependence of observations at the sensors by $c(v_1, v_2, \dots, v_L)$ which is independent of the marginal pdfs so that marginal pdfs can be arbitrary pdfs of continuous random variables and not be limited to Gaussian pdfs. Since the nonparametric copula estimation methods require heavy computations and cannot easily use the knowledge of quantizers, we concentrate on the parametric estimation of $c(v_1, v_2, \dots, v_L)$. Thus, the parametric structure $c(v_1, v_2, \dots, v_L|H_j, \theta_{0j})$ and $p_i(y_i|H_j, \theta_{ij})$ are assumed known under hypothesis H_j , and then we have the joint pdf under hypothesis H_j as follows

$$p(y_1, y_2, \dots, y_L|H_j, \theta_j) = c(F_1(y_1|H_j), F_2(y_2|H_j), \dots, F_L(y_L|H_j)|H_j, \theta_{0j}) \prod_{i=1}^L p_i(y_i|H_j, \theta_{ij}), \quad (9)$$

where $\theta_j \triangleq [\theta_{0j}, \theta_{1j}, \dots, \theta_{Lj}]$ is the parameter vector to be estimated; θ_{0j} is the dependence parameter and $\theta_{1j}, \dots, \theta_{Lj}$ are the marginal parameters under hypothesis $j = 0, 1$. There exist many parametric structures of copula density $c(v_1, v_2, \dots, v_L|H_j, \theta_{0j})$ such as Clayton copula, Gumbel copula, Frank copula and Gauss copula, t copula etc. (see, e.g., [23]). The “best” copula model can be selected by criteria such as Akaike information criterion (AIC), AIC with a correction (AIC_c) and Bayesian information criterion (BIC) etc (see, e.g., [24]). Here, we assume that the copula model has been determined but its parameters are not known.

3.2 Maximum likelihood estimation of unknown prior probabilities and parameters of pdfs with quantized observations

The observation population of the i -th sensor is denoted Y_i , $i = 1, \dots, L$. The observation samples of Y_i may be from H_0 or H_1 . The joint observation population is denoted by $Y \triangleq (Y'_1, \dots, Y'_L)'$ which has the following family of joint pdf:

$$\{p(y_1, \dots, y_L|\theta)\}_{\theta \in \Theta \subseteq \mathbb{R}^k}, \quad (10)$$

where θ is the unknown k -dimensional deterministic parameter vector which may include marginal parameters and dependence parameters. Note that the conditionally joint pdf under hypothesis H_j , $p(y_1, y_2, \dots, y_L|H_j, \theta_j)$ can be constructed by (9) where $\theta_j = [\theta_{0j}, \theta_{1j}, \dots, \theta_{Lj}]$, $j = 0, 1$ are parameter vectors and the prior probabilities of H_0 and H_1 are P_0 and $P_1 = 1 - P_0$ respectively. Thus, we have

$$p(y_1, \dots, y_L|\theta) = P_0 p(y_1, \dots, y_L|H_0, \theta_0) + (1 - P_0) p(y_1, \dots, y_L|H_1, \theta_1), \quad (11)$$

where $\theta \triangleq [P_0, \theta_0, \theta_1]$ is the parameter vector to be estimated. Note that θ_0 and θ_1 themselves are vectors. The true parameter vector (the clairvoyant case) is denoted by θ^* .

Let N independently and identically distributed (i.i.d.) temporal sensor observation samples and joint observation samples be

$$\vec{Y}_i = (Y_{i1}, \dots, Y_{iN}), \quad i = 1, \dots, L; \quad (12)$$

$$\vec{Y} = (\vec{Y}_1', \dots, \vec{Y}_L')'. \quad (13)$$

Moreover, based on the definition of quantizers (1)–(3), we define the quantized sensor observation samples and the joint quantized observation samples as follows

$$\vec{U} \triangleq (\vec{U}_1', \dots, \vec{U}_L')', \quad (14)$$

$$\vec{U}_i \triangleq (U_{i1}, \dots, U_{iN})', \quad i = 1, \dots, L, \quad (15)$$

$$U_{in} \triangleq (U_{in}^1, \dots, U_{in}^{r_i}), \quad n = 1, \dots, N, \quad (16)$$

$$U_{in}^1 \triangleq I_i^1(Y_{in}), \dots, U_{in}^{r_i} \triangleq I_i^{r_i}(Y_{in}), \quad (17)$$

where \vec{U} is the joint quantized observation samples. We denote the quantized observation population by $U \triangleq I(Y|r) = (I_1(Y_1), \dots, I_L(Y_L))'$, we know that U has a *discrete/categorical* distribution. Based on the pdf of Y and quantizers $I(y|r)$, the probability mass function (pmf) of the quantized observation population U is

$$f_U(u_1, u_2, \dots, u_L|\theta) = P_{u_1, u_2, \dots, u_L} \quad \text{for } U = (u_1, u_2, \dots, u_L), \quad (18)$$

where

$$(u_1, u_2, \dots, u_L) \in \mathcal{S}_u = \{(u_1, u_2, \dots, u_L) \in \mathbb{R}^r : \\ u_i \text{ is a } r_i\text{-dimensional binary row vector, } i = 1, \dots, L, r = \sum_{i=1}^L r_i\}, \quad (19)$$

$$P_{u_1, u_2, \dots, u_L} = \int_{\Xi(u_1, u_2, \dots, u_L)} p(y_1, y_2, \dots, y_L|\theta) dy_1 dy_2 \dots dy_L, \quad (20)$$

$$\Xi(u_1, u_2, \dots, u_L) = \{(y_1, y_2, \dots, y_L) : I_1(y_1) = u_1, I_2(y_2) = u_2, \dots, I_L(y_L) = u_L\}. \quad (21)$$

Note that $f_U(u_1, u_2, \dots, u_L|\theta)$ is determined by $p(y_1, y_2, \dots, y_L|\theta)$ and sensor quantizers $I_1(y_1), \dots, I_L(y_L)$.

Thus, the quantized observation population U has a pmf parameter family $\{f_U(u_1, u_2, \dots, u_L|\theta)\}_{\theta \in \Theta \subseteq \mathbb{R}^k}$ which yields the following log likelihood function of quantized samples \vec{U} by (18)-(21):

$$l(\theta|\vec{U}) \triangleq \log \prod_{n=1}^N f_U(U_{1n}, U_{2n}, \dots, U_{Ln}|\theta) \quad (22)$$

$$\begin{aligned}
&= \sum_{n=1}^N \log \int_{\{I_1(y_1)=U_{1n}, \dots, I_L(y_L)=U_{Ln}\}} p(y_1, \dots, y_L | \theta) dy_1 \dots dy_L \\
&= \sum_{m=1}^{2^r} K_m \log f_U(\vec{u}_m | \theta)
\end{aligned} \tag{23}$$

where $K_m = \#\{(U_{1n}, U_{2n}, \dots, U_{Ln}) = \vec{u}_m \in S_u, n = 1, \dots, N\}$, $\sum_{m=1}^{2^r} K_m = N$; S_u is defined by (19); $\#\{\cdot\}$ is the cardinality of the set. The parameter vector θ is estimated by maximizing the log likelihood function (23) or equivalently solving the equation:

$$\frac{\partial}{\partial \theta} l(\theta | \vec{U}) = 0, \tag{24}$$

whose solution is denoted by $\hat{\theta}$. In [21], we have considered the estimation problem in detail and have presented the regularity conditions for $p(y_1, y_2, \dots, y_L | \theta)$ and quantizers $I(y|r)$ that guarantee that $\hat{\theta}$ is asymptotically efficient.

3.3 Maximum likelihood estimation with feedback

As indicated earlier, to improve the detection performance in the distributed detection system, the quantizers are updated and fed back to the sensors for use at the following time. The quantizers $I(y|r)$ defined in (2) used at a given time j , $j = 1, \dots, J$, are known as one group of quantizers. To distinguish different groups, we use superscript (j) and change notations n, N to n_j, N_j respectively, $j = 1, \dots, J$. Assume that, for the j -th group of quantizers $I^{(j)}(y|r)$, N_j joint samples $\{(Y_{1n_j}, \dots, Y_{Ln_j})\}_{n_j=1}^{N_j}$ are observed and the quantized observations are denoted by $\vec{U}^{(j)}$. The corresponding observation population denoted by $U^{(j)}$ whose pmf can be similarly defined by (18) and be denoted by

$$f_U^{(j)}(u_1, u_2, \dots, u_L | \theta), j = 1, \dots, J. \tag{25}$$

Since the samples are temporally independent, we can estimate θ by maximizing the log likelihood function:

$$\begin{aligned}
l(\theta | \vec{U}^{(1)}, \dots, \vec{U}^{(J)}) &= \log \prod_{j=1}^J \prod_{n_j=1}^{N_j} f_U^{(j)}(U_{1n_j}^{(j)}, U_{2n_j}^{(j)}, \dots, U_{Ln_j}^{(j)} | \theta) \\
&= \sum_{j=1}^J \sum_{n_j=1}^{N_j} \log f_U^{(j)}(U_{1n_j}^{(j)}, U_{2n_j}^{(j)}, \dots, U_{Ln_j}^{(j)} | \theta) \\
&= \sum_{j=1}^J \sum_{n_j=1}^{N_j} \log \int_{\{I_1^{(j)}(y_1)=U_{1n_j}^{(j)}, \dots, I_L^{(j)}(y_L)=U_{Ln_j}^{(j)}\}} p(y_1, \dots, y_L | \theta) dy_1 \dots dy_L
\end{aligned} \tag{26}$$

$$= \sum_{j=1}^J \sum_{m=1}^{2^r} K_m^{(j)} \log f_U^{(j)}(\vec{u}_m | \theta) \quad (27)$$

where $K_m^{(j)} = \#\{(U_{1n_j}^{(j)}, U_{2n_j}^{(j)}, \dots, U_{Ln_j}^{(j)}) = \vec{u}_m \in S_u, n_j = 1, \dots, N_j\}$, $\sum_{m=1}^{2^r} K_m^{(j)} = N_j$; S_u is defined by (19); $\#\{\cdot\}$ is the cardinality of the set. Equivalently, we solve the equation:

$$\frac{\partial}{\partial \theta} l(\theta | \vec{U}^{(1)}, \dots, \vec{U}^{(J)}) = 0, \quad (28)$$

whose solution is denoted by $\hat{\theta}_R$. In [21], we have proved that $\hat{\theta}_R$ is an asymptotically efficient estimator with an asymptotic variance equal to the inverse of a convex linear combination of Fisher information matrices based on J groups of different quantizers. These results are summarized in the following Lemma.

Lemma 3.2. *There are J groups of different sensor quantizers $I^{(j)}(y|r)$, $j = 1, \dots, J$. Assume that $p(y_1, y_2, \dots, y_L | \theta)$ and quantizers $I^{(j)}(y|r)$ generate the quantized observations and the quantized pmf $f_U^{(j)}(u_1, u_2, \dots, u_L | \theta)$ defined by (25) satisfies the regularity conditions (C1)–(C7) in [21]. The true parameter vector is denoted by θ^* . Then,*

$$\sqrt{N}(\hat{\theta}_R - \theta^*) \longrightarrow N(0, \mathcal{I}^{-1}(\theta^*; I^{(1)}(\cdot), \dots, I^{(J)}(\cdot))) \quad (29)$$

where $N = \sum_{j=1}^J N_j$, $N_j \rightarrow \infty$, $j = 1, \dots, J$,

$$\mathcal{I}^{-1}(\theta^*; I^{(1)}(\cdot), \dots, I^{(J)}(\cdot)) \triangleq \left(\sum_{j=1}^J \frac{N_j}{N} \mathcal{I}(\theta^*; I^{(j)}(\cdot)) \right)^{-1} \quad (30)$$

$$= N \left(\sum_{j=1}^J N_j \mathcal{I}(\theta^*; I^{(j)}(\cdot)) \right)^{-1}. \quad (31)$$

$\left(\sum_{j=1}^J N_j \mathcal{I}(\theta^*; I^{(j)}(\cdot)) \right)^{-1}$ is the Cramér-Rao lower bound for N quantized observations, where $\mathcal{I}(\theta^*; I^{(j)}(\cdot))$ is the Fisher information matrix for one quantized sample of $U^{(j)}$. That is, $\hat{\theta}_R$ is a consistent and asymptotically efficient estimator of θ^* .

4 Distributed Detection System Design Using MLE with Feedback

When the prior probabilities and conditional pdfs are unknown, the basic idea of the distributed detection system design is as follows. We begin with an initial set of quantizers at the sensors and send quantized observations to the fusion center. The fusion center starts with an initial fusion rule. Based on the received quantized observations, the fusion center computes the MLE for the unknown parameters, and obtains updated quantizers and the fusion rule. Updated quantizers are fed back to the sensors and are used to quantize the next set of observations. This iterative process is continued several times to continually improve the

parameter estimates and thereby improving detection performance. In summary, based on the MLE with feedback and Algorithm 1 in [17] which is a near-optimal iterative algorithm and can simultaneously design the quantizers and the fusion rule when the prior probabilities and the conditionally dependent pdfs are known, we have the following algorithm.

Algorithm 4.1 (Distributed detection system design based on MLE with feedback).

Step 1: Initialize L quantizers and the fusion rule at first stage respectively, for $i = 1, \dots, L$,

$$I_i^{(1)}(y_{im_i}) = \mathbf{0}/\mathbf{1}, \text{ for } m_i = 1, \dots, M_i, \quad (32)$$

$$F^{(1)}(\vec{u}_j) = 0/1, \text{ for } \vec{u}_j \in S_u, j = 1, \dots, 2^L, \quad (33)$$

where $\mathbf{0}/\mathbf{1}$ is r_i dimensional 0/1 vector; the measurement space of the i -th sensor is discretized to M_i regions; S_u are defined by (19). N_1 samples are sequentially observed and quantized to binary samples which are sent to the fusion center. Let $t = 1$ and $J = 1$, go to next step.

Step 2: Estimate parameter θ at the t -th stage: the MLE with feedback $\hat{\theta}_R$ is computed based on all previous quantized observations by maximizing (27). Thus, we have $p(y_1, y_2, \dots, y_L | \hat{\theta}_R)$, go to next step.

Step 3: Design sensor quantizers and the fusion rule at the t -th stage: Based on $p(y_1, y_2, \dots, y_L | \hat{\theta}_R)$ and Algorithm 1 in [17], sensor quantizers and the fusion rule are iteratively searched for better detection performance until the termination criterion of Algorithm 1 in [17] is satisfied. Thus, we have the quantizers $I_i^{(t)}(\cdot), i = 1, \dots, L$ and the fusion rule $F^{(t)}(\cdot)$, go to next step.

Step 4: Feedback: If $t \leq T$ (T is the upper bound on the number of feedbacks), let $t = t + 1$ and set $J = t$. The fusion center sends the current quantizers $I_i^{(t-1)}(\cdot), i = 1, \dots, L$ to the sensors, and then the sensors sequentially observe N_t quantized samples based on the current quantizers and send them to the fusion center. Go to step 2. If $t > T$, stop and the last quantizers are transmitted to the sensors.

In feedback step 4, communication can be reduced by only transmitting the changes of the quantizers between two iteration steps to sensors.

5 Numerical Examples

Let us consider a binary hypothesis testing problem with two sensors

$$\begin{aligned} H_0 : Y_1 &\sim \text{Gamma}(3, 4) \quad Y_2 \sim \text{Gamma}(5, 4), \\ c(v_1, v_2 | H_0) &\equiv 1 \quad (\text{independent copula density}), \end{aligned} \quad (34)$$

$$H_1 : Y_1 \sim \text{Gamma}(5, 4) \quad Y_2 \sim \text{Gamma}(7, 4),$$

$$c(v_1, v_2|H_1, \theta_1) = (1 + \theta_1)v_1^{-1-\theta_1}v_2^{-1-\theta_1}(-1 + v_1^{-\theta_1} + v_2^{-\theta_1})^{-2-1/\theta_1}, \theta_1 \in [-1, \infty) \setminus \{0\}, (35)$$

where the prior probability P_1 and dependence parameter θ_1 of hypothesis H_1 are unknown and are required to be estimated. We denote by $\theta \triangleq [P_1, \theta_1]$. Here, we assume that the joint pdf under H_0 is independent and marginal pdfs are known. $c(v_1, v_2|H_1, \theta_1)$ is the Clayton copula density (see e.g. [23]) which is a frequently used copula to model dependence. The actual value of θ (ground truth) is that $P_1 = 0.2$ and $\theta_1 = [0.5109, 1.0759, 2.1316]$ which corresponds to Spearman's dependence measure $\rho = [0.3, 0.5, 0.7]$ ¹. There is a one to one relationship between θ_1 and ρ (see [23]).

The initial values of the quantizers chosen are $I_1(y_1) = I[3y_1 - 60]$, $I_2(y_2) = I[-3y_2 + 60]$, and the initial fusion rule used is the OR fusion rule. For numerical computation, we take a discretization step-size $\Delta = 0.5$, $y_i \in [0, 60]$. We denote the probability of a false alarm and the probability of detection by P_f and P_d respectively.

In Figure 2, RMSEs of the MLE with feedback for the prior probability P_1 are plotted based on 5000 Monte Carlo runs. Three dependence cases with Spearman's $\rho = [0.3, 0.5, 0.7]$ are considered. The feedback times increase from $J = 2$ to $J = 10$ and each estimation step is done after receiving 100 quantized observations so that the number of observations becomes from $J \times 100 = 200$ to 1000. The cost parameters $C_{00} = C_{11} = 0$ and $C_{10} = 2$, $C_{01} = 1$ are used in cost function Eq. (5). Similarly, RMSEs of the MLE with feedback for the dependence parameter θ_1 are plotted in Figure 3.

In Figures 4–6, to evaluate the detection performance of Algorithm 4.1, the average ROC curves based on 500 Monte Carlo runs are compared for the following three cases: 1) Algorithm 4.1 based on MLE with feedback, where “Algorithm 4.1–10*100” means the number of feedbacks $J = 10$ and the number of observations in each estimation step $N_j = 100$, $j = 1, \dots, J$. 2) Assume independence under H_1 so that the joint density is the product of marginals and with known prior probabilities. 3) The case with known parameter values (the clairvoyant case). In each Monte Carlo run, 400 test observations from H_1 and 1600 test observations from H_0 are generated so that P_f and P_d can be computed respectively. To plot several points on the ROCs, the cost parameters $C_{00} = C_{11} = 0$ and $C_{10} = 2$, $C_{01} = 2 \times [0.6, 0.5, 0.4, 0.3, 0.25, 0.2, 0.15, 0.1, 0.05, 0.02]$ are used in cost function Eq. (5).

From Figures 2–6, we have the following observations:

1. From Figures 2 and 3, RMSEs decrease as the number of observations increases. However, the rate of decay becomes slow, especially in Figure 3. Thus, to improve performance at a later stage, more feedback times and samples are required.
2. RMSE of the parameter P_1 in the case of $\rho = 0.7$ is the smallest among the three cases in Figure 2. However, RMSE of the parameter θ_1 in the case of $\rho = 0.7$ is largest in Figure 3. The reason may be that the estimated pdf in this case is closer to the actual pdf. In addition, from Figure 3, RMSE of

¹Spearman's $|\rho| \leq 1$ is a commonly used dependence measure (see [23]).

θ_1 for $\rho = 0.3$ is less than those for $\rho = 0.5$ and $\rho = 0.7$. The reason may be that the value of θ_1 corresponding to $\rho = 0.3$ is less than those corresponding to $\rho = 0.5$ and $\rho = 0.7$.

3. From Figures 4–6, the new detection Algorithm 4.1 based on MLE with feedback is much better than the usual detection method based on the assumption of independent observations. For fixed number of observations in each estimation step, the better performance can be obtained by increasing the number of feedbacks from 5 to 10. For fixed the number of feedbacks 10, better performance can be obtained by increasing the number of observations in each estimation step, especially for the case of the larger dependence parameter $\rho = 0.7$.

6 Conclusion

In this paper, distributed Bayesian detection problems with unknown prior probabilities of hypotheses and unknown conditional pdfs have been considered. The conditional joint pdf was represented in a parametric form by using the copula framework. The unknown parameters included dependence parameters and marginal parameters. MLE with feedback based on quantized data has been proposed to estimate the unknown parameters. Its asymptotic efficiency can be guaranteed by employing the result that we have developed in [21] with an asymptotic variance which is equal to the inverse of a convex linear combination of Fisher information matrices based on J groups of different feedback quantizers. These estimates were iteratively used to refine the quantizers and the fusion rule to improve distributed detection performance by using feedback. Numerical examples show that the new detection method based on MLE with feedback is much better than the usual detection method based on the assumption of conditionally independent observations. Better detection performance can be obtained by increasing the number of feedbacks and the number of observations in each estimation step.

Future work will involve distributed detection and distributed location estimation of non-ideal systems where sensor outputs are to be communicated through noisy, possibly coupled channels to the fusion center.

Acknowledgment

We would like to thank Hao He and Arun Subramanian for their suggestions on simulations of this paper.

References

- [1] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1997.
- [2] R. Vismanathan and P. K. Varshney, “Distributed detection with multiple sensors: Part I-fundamentals,” *Proceeding of IEEE*, vol. 85, pp. 54–63, 1997.

- [3] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II advanced topics," *Proceedings of the IEEE*, vol. 85, pp. 64–C79, 1997.
- [4] Y. Zhu, J. Zhou, X. Shen, E. Song, and Y. Luo, *Networked Multisensor Decision and Estimation Fusion: Based on Advanced Mathematical Methods*. CRC Press, 2012.
- [5] V. V. Veeravalli and P. K. Varshney, "Distributed inference in wireless sensor networks," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 370, pp. 100–117, January 2012.
- [6] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Transaction on Aerospace and Electronic Systems*, vol. 17, no. 4, pp. 501–510, 1981.
- [7] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Transaction on Aerospace and Electronic Systems*, vol. 22, pp. 98–101, January 1986.
- [8] B. Chen, L. Tong, and P. K. Varshney, "Channel-aware distributed detection in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 16–26, 2006.
- [9] B. Chen and P. Willett, "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 693–699, 2005.
- [10] H. Chen, B. Chen, and P. K. Varshney, "Further results on the optimality of likelihood ratio quantizer for distributed detection in nonideal channels," *IEEE Transactions on Information Theory*, vol. 55, pp. 828–832, February 2009.
- [11] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Transactions on Automatic Control*, vol. 30, pp. 440–446, 1985.
- [12] R. S. Blum and S. A. Kassam, "Optimum-distributed detection of weak signals in dependent sensors," *IEEE Transactions on Information Theory*, vol. 36, pp. 1066–1079, 1992.
- [13] P. N. Chen and A. Papamarcou, "Likelihood ratio partitions for distributed signal detection in correlated gaussian noise," *Proceedings of IEEE International symposium Information Theory*, p. 118, October 1995.
- [14] P. Willett, P. F. Swaszek, and R. S. Blum, "The good, bad and ugly: Distributed detection of a known signal in dependent Gaussian noise," *IEEE Transactions on Signal Processing*, vol. 48, pp. 3266–3279, December 2000.
- [15] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "A distributed M-ary hypothesis testing problem with correlated observations," *IEEE Transactions on Automatic Control*, vol. 37, pp. 1042–1046, July 1992.

- [16] Y. Zhu, R. S. Blum, Z.-Q. Luo, and K. M. Wong, “Unexpected properties and optimum-distributed sensor detectors for dependent observation cases,” *IEEE Transactions on Automatic Control*, vol. 45, pp. 62–72, January 2000.
- [17] X. Shen, Y. Zhu, L. He, and Z. You, “A near-optimal iterative algorithm via alternately optimizing sensor and fusion rules in distributed decision systems,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, pp. 2514–2529, October 2011.
- [18] H. Chen, B. Chen, and P. K. Varshney, “A new framework for distributed detection with conditionally dependent observations,” *IEEE Transactions on Signal Processing*, vol. 60, pp. 1409–1419, March 2012.
- [19] A. Sundaresan, P. K. Varshney, and N. S. V. Rao, “Copula-based fusion of correlated decisions,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, pp. 454–471, January 2011.
- [20] S. G. Iyengar, P. K. Varshney, and T. Damarla, “A parametric copula-based framework for hypothesis testing using heterogeneous data,” *IEEE Transactions on Signal Processing*, vol. 59, pp. 2308–2319, May 2011.
- [21] X. Shen, P. K. Varshney, and Y. Zhu, “Robust distributed maximum likelihood estimation with quantized data.” <http://arxiv.org/abs/1208.4161>, August 2012.
- [22] A. Sklar, “Fonctions de répartition à n dimensions et leurs marges,” (*French*) *Publ. Inst. Statist. Univ. Paris.*, vol. 8, pp. 229–231, 1959.
- [23] R. B. Nelsen, *An Introduction to Copulas*. Springer-Verlag, New York, 1999.
- [24] K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer-Verlag, second ed., 2002.

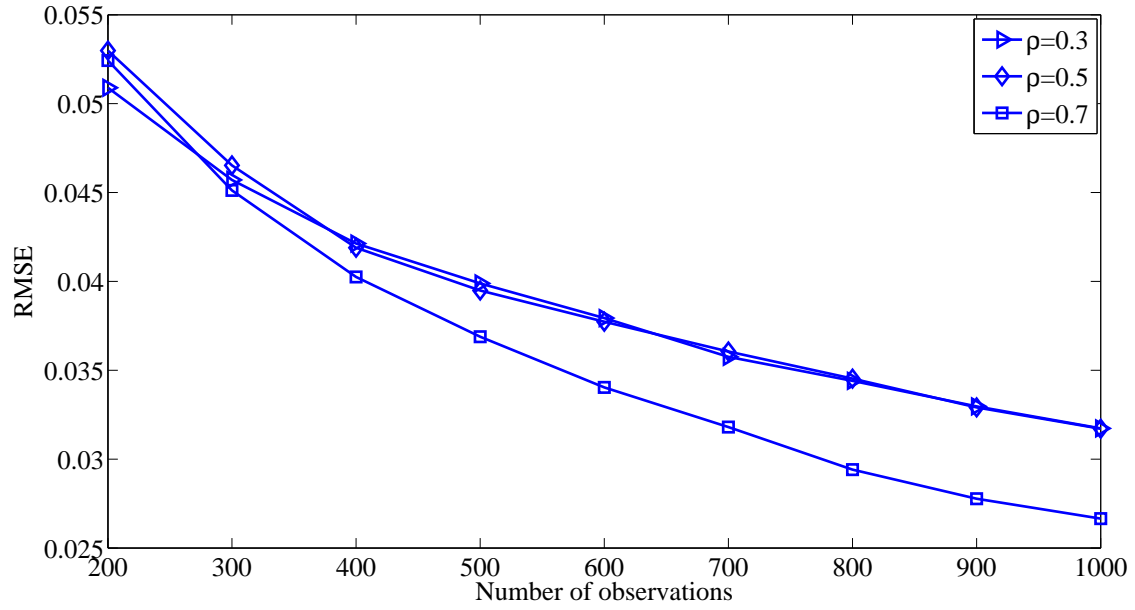


Figure 2: RMSEs of the prior probability P_1 based on 5000 Monte Carlo runs for the cases of the dependence measure Spearman's $\rho = [0.3, 0.5, 0.7]$.

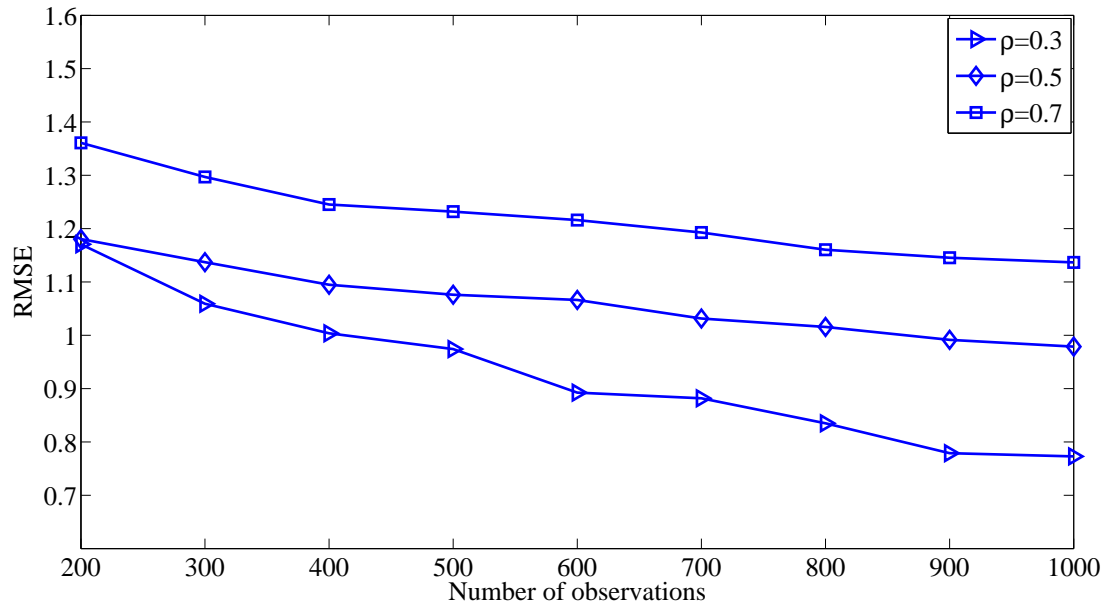


Figure 3: RMSEs of the dependence parameter θ_1 based on 5000 Monte Carlo runs for the cases of the dependence measure Spearman's $\rho = [0.3, 0.5, 0.7]$

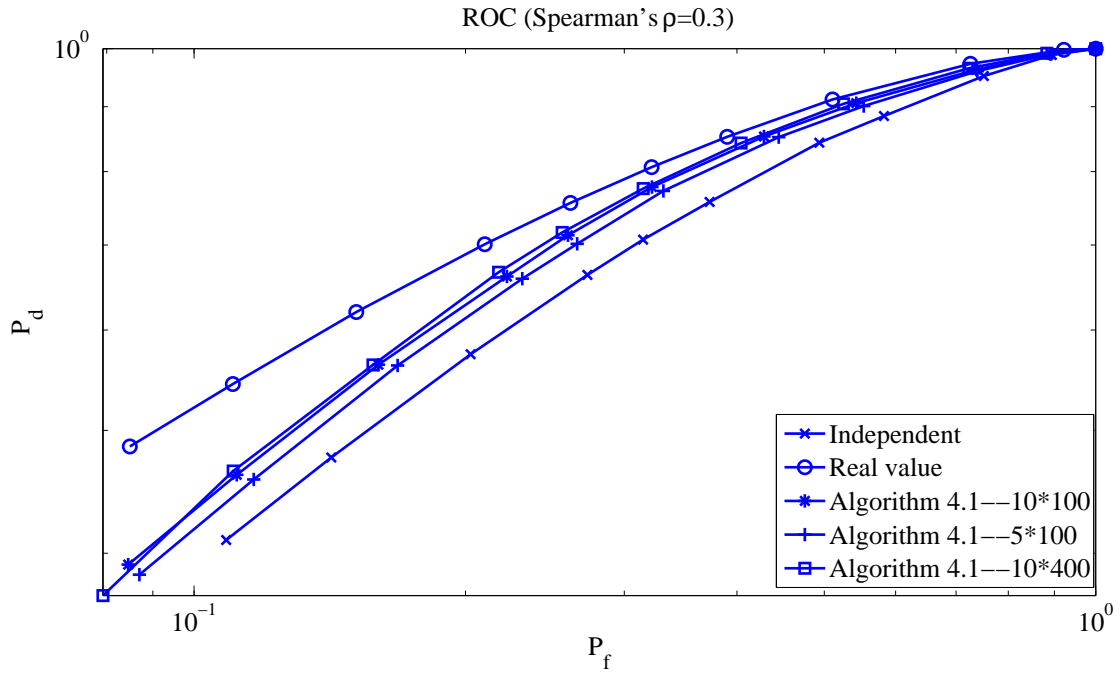


Figure 4: Comparison of ROCs for three cases with Spearman's $\rho = 0.3$.

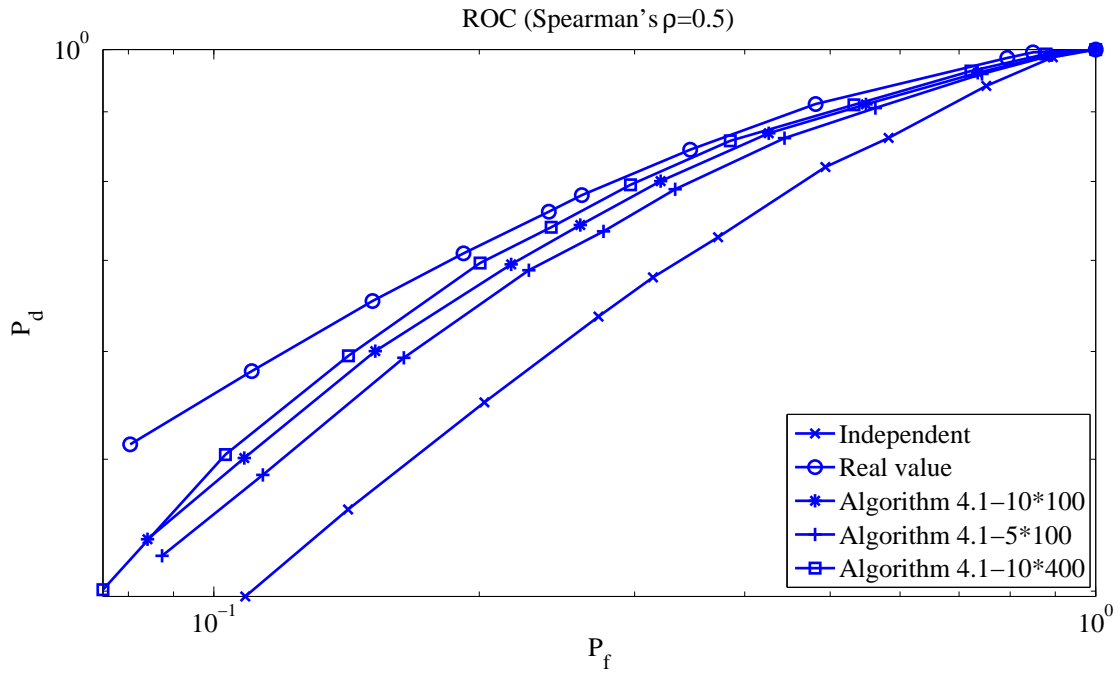


Figure 5: Comparison of ROCs for three cases with Spearman's $\rho = 0.5$.

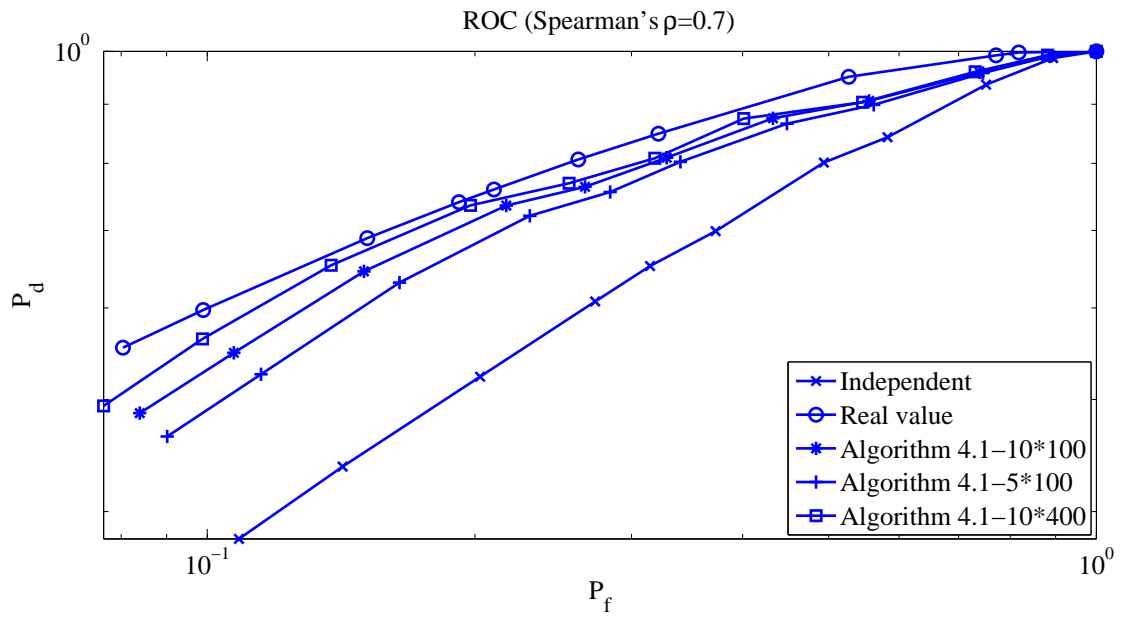


Figure 6: Comparison of ROCs for three cases with Spearman's $\rho = 0.7$.